REMARKS

In the non-final Office Action, dated February 8, 2008, the Examiner rejects

claims 20-30 and 34 under 35 U.S.C. § 101 as allegedly being directed to non-statutory

subject matter; rejects claims 1, 2, 5, 6, 8, 11, 13, 16, 17, 20, 24, and 26-28 under 35

U.S.C. § 102(a) as allegedly being anticipated by Schleimer et al., "Winnowing: Local

Algorithms for Document Fingerprinting," published June 9, 2003 (hereinafter

"SCHLEIMER"); rejects claim 20 under 35 U.S.C. § 102(b) as allegedly being

anticipated by U.S. Patent No. 5,745,900 B1 to Burrows (hereinafter "BURROWS");

rejects claims 1-2, 5-8, 10, 13, 16-17, 24-29, 31, 32, and 34 under 35 U.S.C. § 103(a) as

allegedly being unpatentable over BURROWS in view of U.S. Published Patent

Application No. 2002/0133499 A1 to Ward et al. (hereinafter "WARD"); rejects claims

3, 4, 11, 12, 14 and 15 under 35 U.S.C. § 103(a) as allegedly being unpatentable over

BURROWS and WARD and further in view of U.S. Patent No. 6,230,155 B1 to Broder

et al. (hereinafter "BRODER"); rejects claims 22 and 23 under 35 U.S.C. § 103(a) as

allegedly being unpatentable over BURROWS in view of Charikar, "Similarity

Estimation Techniques from Rounding Algorithms", published May 19, 2002 (hereinafter

"CHARIKAR"); rejects claims 9, 18-19, 30, and 33 under 35 U.S.C. § 103(a) as

allegedly being unpatentable over BURROWS and WARD and further in view of

Official Notice; and rejects claim 21 under 35 U.S.C. § 103(a) as allegedly being

unpatentable over BURROWS in view of Official Notice. Applicant respectfully

traverses these rejections.[1]

---

[1] As Applicant's remarks with respect to the Examiner's rejections are sufficient to overcome these rejections, Applicant's silence as to assertions by the Examiner in the Office Action or certain requirements that may be applicable to such rejections (e.g., whether a reference constitutes prior art (reasons to modify a reference or to combine references, assertions as to dependent claims, etc.) is not a concession by Applicant that such assertions are accurate or such requirements have been met, and Applicant reserves the right to analyze and dispute such assertions/requirements in the future.

By way of the present amendment, Applicant amends claims 1, 12, 20, 26, 31, and 34 to improve form. Support for the amendments to claims 1, 20, 26, and 34 can be found, for example, in paragraph [0038]. Support for the amendments to claims 20, 26, and 34 can be found, for example, in Fig. 3, items 330 and 320. No new matter has been added by way of the present amendment. Claims 1-34 remain pending.

Claims 20-30 and 34 stand rejected under 35 U.S.C. § 101 as being allegedly directed to non-statutory subject matter. Applicant respectfully traverses this rejection.

The Examiner alleged that the recitation "computer-implemented device" in previously amended claims 20-30 and 34 can be reasonably interpreted as software and thus these claims allegedly remain non-statutory under 35 U.S.C. § 101 (Office Action, p. 2). Without necessarily agreeing with the Examiner, in order to address the Examiner's concerns and expedite prosecution, Applicant has amended claims 20, 26, and 34 to recite a memory and a processor. Amended claims 20, 26, and 34 are clearly directed to the statutory class of machine and these hardware elements clearly obviate the Examiner's allegation of elements that can be interpreted as only software.

For at least the foregoing reasons, Applicant requests that the rejection of claims 20-30 and 34 under 35 U.S.C. § 101 be reconsidered and withdrawn.

Claims 1, 2, 5, 6, 8, 11, 13, 16, 17, 20, 24, and 26-28 stand rejected under 35 U.S.C. § 102(a) as allegedly being anticipated by SCHLEIMER. Applicant respectfully traverses this rejection.

A proper rejection under 35 U.S.C. § 102 requires that a reference teach every aspect of the claimed invention. Any feature not directly taught must be inherently present. See M.P.E.P. § 2131. SCHLEIMER does not disclose the combination of

features recited in Applicant's claims 1, 2, 5, 6, 8, 11, 13, 16, 17, 20, 24, and 26-28.

For example, amended independent claim 1 is directed to a method for generating a representation of a document. The method includes obtaining a plurality of overlapping blocks by sampling the document, choosing a subset of the plurality of overlapping blocks, where the subset is less than an entirety of the plurality of overlapping blocks, and compacting the subset of the plurality of overlapping blocks to obtain the representation of the document. SCHLEIMER does not disclose or suggest this combination of features.

For example, SCHLEIMER does not disclose or suggest choosing a subset of the plurality of overlapping blocks, where the subset is less than an entirety of the plurality of overlapping blocks, as recited in amended claim 1. The Examiner relies on Section 3 and Fig. 2(e) of SCHLEIMER for allegedly disclosing this feature (Office Action, p. 9). Applicant disagrees with the Examiner's interpretation of SCHLEIMER.

Section 3 of SCHLEIMER describes the winnowing algorithm for selecting fingerprints from hashes of k-grams. In the example of Fig. 2(a)-2(g) of SCHLEIMER, a set of overlapping 5-grams is generated from a set of text (Fig. 2(c) of SCHLEIMER). A sequence of hashes is then generated from the sequence of 5-grams (Fig. 2(d) of SCHLEIMER). Next, a sequence of overlapping windows of the hashes is generated (Fig. 2(e) of SCHLEIMER), followed by the selection of the minimum hash from each window (Fig. 2(f) of SCHLEIMER). This sequence of minimum hashes becomes the fingerprint of the text. This section of SCHLEIMER does not disclose or suggest choosing a subset of a plurality of overlapping blocks that is obtained by sampling a document, where the subset is less than the entirety of the plurality of overlapping blocks,

as recited in amended claim 1. Instead, SCHLEIMER appears to hash all of the 5-grams derived from the text (see Figs. 2(c) and 2(d)), and does not choose a subset of the 5-grams.

In the Response to Arguments section of the Office Action, the Examiner did not address the arguments submitted by Applicant. Applicant argued that SCHLEIMER does not disclose or suggest choosing a subset of the plurality of overlapping blocks, as recited in previously presented claim 1. The Examiner did not indicate on which section or elements of SCHLEIMER the Examiner was relying on for allegedly disclosing choosing a subset of plurality of overlapping blocks, as recited in claim 1. Instead, the Examiner agreed with Applicant that SCHLEIMER does not choose a subset of 5-grams, as would have been required by previously presented claim 1 based on the Examiner's interpretation of SCHLEIMER (Office Action, p. 3). If this rejection is maintained, Applicant respectfully requests that the Examiner indicate which section of SCHLEIMER the Examiner is relying on for allegedly disclosing choosing a subset of the plurality of overlapping blocks, where the subset is less than the entirety of the plurality of overlapping blocks, as recited in amended claim 1.

Instead, the Examiner provided the following arguments. The Examiner alleges that the term "overlapping" is not defined in the specification (Office Action, p. 3). Applicant submits that there is no requirement for defining common words and terms, and that the intended meaning of the term "overlapping" would be clear to one of ordinary skill in the art based on Fig. 6 and paragraph [0035] of the present application. Nevertheless, Applicant submits that this allegation by the Examiner does not address choosing a subset of the plurality of overlapping blocks, where the subset is less than the

entirety of the plurality of overlapping blocks, as recited in amended claim 1.

The Examiner alleges that the phrase "obtain a plurality of overlapping blocks" is the intended use of "sampling the document" and is not given any patentable weight (Office Action, p. 3). While not agreeing with the Examiner, Applicant has amended claim 1 to recite "obtaining a plurality of overlapping blocks by sampling the document" to address the Examiner's concern and in order to expedite prosecution. Nevertheless, Applicant submits that this allegation by the Examiner does not address choosing a subset of the plurality of overlapping blocks, where the subset is less than the entirety of the plurality of overlapping blocks, as recited in amended claim 1.

In response to Applicant's argument that the windows of hashes depicted in Fig. 2(e) of SCHLEIMER are not obtained by sampling the document, the Examiner argues that winnowing is a sampling process and that since SCHLEIMER discloses winnowing of the hashes, and since the hashes are based on the document itself, SCHLEIMER allegedly discloses sampling the document (Office Action, p. 3). Once again, Applicant submits that this allegation by the Examiner does not address choosing a subset of the plurality of overlapping blocks, where the subset is less than the entirety of the plurality of overlapping blocks, as recited in amended claim 1.

Furthermore, as Applicant previously pointed out (see p. 13 of the Amendment filed November 21, 2007) that if the Examiner relies on the windows of hashes of Fig. 2(e) of SCHLEIMER as allegedly corresponding to "obtaining the plurality of overlapping blocks by sampling a document," as recited in amended claim 1, the Examiner cannot also rely on Fig. 2(e) as allegedly corresponding to "choosing a subset of the plurality of overlapping blocks, where the subset is less than an entirety of the

plurality of overlapping blocks," as also recited in amended claim 1. On the other hand, if the Examiner relies on the sequence of 5-grams of Fig. 2(c) of SCHLEIMER as allegedly corresponding to "obtaining the <u>plurality of overlapping blocks</u> by sampling a document," as recited in amended claim 1, the Examiner cannot rely on the windows of hashes Fig. 2(e) of SCHLEIMER as allegedly corresponding to "choosing a subset of the <u>plurality of overlapping blocks</u>, where the subset is less than an entirety of the plurality of overlapping blocks," as also recited in amended claim 1, since the 5-grams and windows of hashes of SCHLEIMER are not even remotely equivalent.

Therefore, no matter how Figs. 2(a) – 2(g) of SCHLEIMER are interpreted, they simply cannot be construed as disclosing or suggesting both "obtaining the <u>plurality of overlapping blocks</u> by sampling a document," as recited in amended claim 1, and "choosing a subset of the <u>plurality of overlapping blocks</u>, where the subset is less than an entirety of the plurality of overlapping blocks," as also recited in amended claim 1. Moreover, as pointed out above, SCHLEIMER does not disclose or suggest choosing a subset of any kind, whether it be the set of 5-grams, the sequence of hashes, the set of windows of hashes, or the fingerprints depicted in Figs. 2(a) – 2(g) of SCHLEIMER, <u>where the subset is less than the entirety of the set</u>.

For at least the foregoing reasons, Applicant submits that claim 1 is not anticipated by SCHLEIMER. Accordingly, Applicant respectfully requests that the rejection of claim 1 under 35 U.S.C. § 102(a) based on SCHLEIMER be reconsidered and withdrawn.

Claims 2, 5, 6, 8, and 11 depend from claim 1. Therefore these claims are not anticipated by SCHLEIMER for at least the reasons set forth above with respect to claim

1. Accordingly, Applicant respectfully requests that the rejection of claims 2, 5, 6, 8, and 11 under 35 U.S.C. § 102(a) based on SCHLEIMER be reconsidered and withdrawn. Moreover, these claims are not anticipated by SCHLEIMER for reasons of their own.

For example, claim 2 recites that compacting the subset of the plurality of overlapping blocks includes setting bits in the representation of the document based on the subset of the plurality of overlapping blocks. SCHLEIMER does not disclose or suggest this feature. The Examiner relies on Fig. 2(g) of SCHLEIMER as allegedly disclosing this feature (Office Action, p. 5). Applicant disagrees with the Examiner's interpretation of SCHLEIMER.

Fig. 2(g) of SCHLEIMER discloses pairing the fingerprint generated by selecting the minimum hash from each hash window of Fig. 2(e) with a number indicating the position of this hash in the sequence of hashes generated from the 5-grams (as shown in Fig. 2(d) of SCHLEIMER). The Examiner relies on the fingerprint of Fig. 2(g) of SCHLEIMER as allegedly corresponding to the representation of the document, as recited in claim 1. SCHLEIMER does not disclose or suggest setting the bits of the fingerprint based on the subset of the windows of hashes, as would be required by claim 2, based on the Examiner's interpretation of SCHLEIMER. In fact, SCHLEIMER does not disclose or suggest setting the bits of the fingerprint at all. SCHLEIMER only discloses pairing each hash of the fingerprint with a number indicating the position of the hash. Therefore, SCHLEIMER cannot disclose or suggest that compacting the subset of the plurality of overlapping blocks includes setting bits in the representation of the document based on the subset of the plurality of overlapping blocks, as recited in claim 2.

In the Response to Arguments section of the Office Action, the Examiner argues

that the act of placing or even selecting the hashes of SCHLEIMER involves storage in memory, which in turn involves the setting of bits (Office Action, p. 4). Applicant disagrees with the Examiner's allegation. Claim 2 does not simply recite that compacting the subset of the plurality of overlapping blocks includes <u>setting bits</u>. Claim 2 recites that compacting the subset of the plurality of overlapping blocks includes <u>setting bits in the representation of the document based on the subset of the plurality of overlapping blocks</u>. While the act of storing something in memory does involve setting bits, it involves setting bits at the location where it is to be stored. The setting of bits by storing something in memory is unrelated to the setting of bits <u>based on the subset of the plurality of overlapping blocks</u>. Therefore, the Examiner's allegation lacks merit.

For at least these additional reasons, Applicant submits that claim 2 is not anticipated by SCHLEIMER.

Independent claim 13 is directed to a method for generating a representation of a document. The method includes sampling the document to obtain a plurality of overlapping samples, selecting a predetermined number of the plurality of overlapping samples as those of the samples corresponding to a predetermined number of smallest samples or a predetermined number of largest samples, and setting bits in the representation of the document based on the selected predetermined number of the samples. SCHLEIMER does not disclose or suggest this combination of features.

The Examiner did not address the features of claim 13 in the Office Action. Instead, the Examiner addressed the features of claim 1 (Office Action, p. 5). <u>Claim 1 does not, for example, recite selecting a predetermined number of the plurality of overlapping samples as those of the samples corresponding to a predetermined number of</u>

smallest samples or a predetermined number of largest samples, as recited in claim 13.
The Examiner does not address this feature. Thus, a proper case of anticipation has not been established with respect to claim 13. If this rejection is maintained, Applicant respectfully requests that the Examiner specifically address the features recited in claim 13.

Applicant further submits that SCHLEIMER does not, for example, disclose or suggest setting bits in the representation of the document based on the selected predetermined number of the samples, as recited in claim 13. This feature is similar to the feature present in claim 2. Therefore, the arguments presented (yet possibly of different scope than) above with respect to claim 2 are applicable here, which was addressed above.

For at least these reasons and for reasons similar to reasons given above with respect to claim 2, Applicant submits that claim 13 is not anticipated by SCHLEIMER. Accordingly, Applicant respectfully requests that the rejection of claim 13 under 35 U.S.C. § 102(a) based on SCHLEIMER be reconsidered and withdrawn.

Claims 16 and 17 depend from claim 13. Therefore, these claims are not anticipated by SCHLEIMER for at least the reasons set forth above with respect to claim 13. Accordingly, Applicant respectfully requests that the rejection of claims 16 and 17 under 35 U.S.C. § 102(a) based on SCHLEIMER be reconsidered and withdrawn.

Amended independent claims 20 and 26 recite features which are similar to, yet possibly of different scope than, the features recited above with respect to claim 1. Therefore, these claims are not anticipated by SCHLEIMER for at least reasons similar to reasons set forth above with respect to claim 1. Accordingly, Applicant respectfully

requests that the rejection of claims 20 and 26 under 35 U.S.C. § 102(a) based on

SCHLEIMER be reconsidered and withdrawn.

Claim 24 depends from claim 20. Therefore, this claim is not anticipated by

SCHLEIMER for at least the reasons set forth above with respect to claim 20.

Accordingly, Applicant respectfully requests that the rejection of claim 24 under 35

U.S.C. § 102(a) based on SCHLEIMER be reconsidered and withdrawn.

Claims 27 and 28 depend from claim 26. Therefore, these claims are not

anticipated by SCHLEIMER for at least the reasons set forth above with respect to claim

26. Accordingly, Applicant respectfully requests that the rejection of claims 27 and 28

under 35 U.S.C. § 102(a) based on SCHLEIMER be reconsidered and withdrawn.

Claim 20 stands rejected under 35 U.S.C. § 102(b) as allegedly being anticipated

by BURROWS. Applicant respectfully traverses this rejection.

Amended independent claim 20 is directed to a computer-implemented device.

The device includes a memory to store instructions for implementing a fingerprint

creation component to generate a fingerprint of a predetermined length for an input

document, the fingerprint generated by sampling the input document to obtain samples,

choosing a subset of the samples, and generating the fingerprint from the subset of the

samples by compacting the subset of the samples, and a similarity detection component to

compare pairs of fingerprints to determine whether the pairs of fingerprints correspond to

near-duplicate documents; and a processor to execute the instructions in the memory.

BURROWS does not disclose or suggest this combination of features.

For example, BURROWS does not disclose or suggest a fingerprint creation

component to generate a fingerprint of a predetermined length for an input document, the

fingerprint generated by sampling the input document to obtain samples, choosing a

subset of the samples, and generating the fingerprint from the subset of the samples by

compacting the subset of the samples, as recited in claim 20. The Examiner relies on

Figs. 4 and 5 of BURROWS for allegedly disclosing this feature (Office Action, p. 12).

Applicant disagrees with the Examiner's interpretation of BURROWS.

Fig. 4 of BURROWS, which is described in col. 7, line 41 to col. 9, line 32,

depicts a block diagram of content attributes generated by the search engine. Fig. 4

shows portions of a page, labeled as 230, 240, 250, and 260, which are detected and

encoded by a parsing module. The parsing module generates attribute values for entire

pages, portions of a page, fields, or individual words and the parser stores these attribute

values as searchable metawords. The fingerprint 255 of Fig. 4 can be one of these

metawords. BURROWS specifically discloses that fingerprint 255 can be produced by

applying one-way polynomial functions to the digitized content of the document (col. 8,

lines 16-23 of BURROWS). Thus, this section of BURROWS does not disclose or

suggests a fingerprint creation component to generate a fingerprint of a predetermined

length for an input document, the fingerprint generated by sampling the input document

to obtain samples, choosing a subset of the samples, and generating the fingerprint from

the subset of the samples by compacting the subset of the samples, as recited in claim 20.

Fig. 5 of BURROWS, described in col. 9, lines 33-41, shows a view of the words

and metawords produced by the parsing module. The parsing module produces a

sequence of pairs in a collating order according to the location of the words of various

pages.

Fig. 5 of BURROWS does not disclose or suggest anything about a fingerprint

creation component. In fact, BURROWS merely discloses that a fingerprint 255 consists

of an integer value generated by applying a one-way polynomial, as described above.

Therefore, this figure of BURROWS cannot disclose or suggest a fingerprint creation

component to generate a fingerprint of a predetermined length for an input document, the

fingerprint generated by sampling the input document to obtain samples, choosing a

subset of the samples, and generating the fingerprint from the subset of the samples by

compacting the subset of the samples, as recited in claim 20.

In the Response to Arguments section of the Office Action, the Examiner quotes

part of the Summary of Invention of BURROWS as (Office Action, p. 5):

> The invention provides a computer implemented method for indexing duplicate information stored as records having different unique addresses in a database. The method generates a fingerprint for each record. The fingerprint is a singular value derived from all of the information of the record according to a predetermined combination of the information of the record.
> (emphasis added by the Examiner)

This section of BURROWS merely discloses that a fingerprint is generated for each

record. As pointed out above, the fingerprint in BURROWS is generated by applying a

one-way polynomial function to the digitized content of a document. This cannot be

reasonably construed as being equivalent to a fingerprint generated by sampling the input

document to obtain samples, choosing a subset of the samples, and generating the

fingerprint from the subset of the samples by compacting the subset of the samples, as

recited in claim 20. The Examiner alleges that Fig. 5 of BURROWS discloses

compacting a subset of samples, where the samples were obtained by sampling a

document, to generate a fingerprint (Office Action, p. 5). However, Fig. 5 of

BURROWS merely shows a parsing module which parses a retrieved web page in

preparation for indexing (see items 200, 30, and 40 in Fig. 2 of BURROWS). The

parsing module 30 and Fig. 5 of BURROWS have nothing to do with generating the

fingerprint 255 shown in Fig. 4 of BURROWS. If this rejection is maintained, Applicant respectfully request that the Examiner point out which section of BURROWS relates Fig. 5 and the parsing module 30 to the generation of fingerprint 255 shown in Fig. 4 of BURROWS. BURROWS does not disclose or suggest a fingerprint creation component to generate a fingerprint of a predetermined length for an input document, the fingerprint generated by sampling the input document to obtain samples, choosing a subset of the samples, and generating the fingerprint from the subset of the samples by compacting the subset of the samples, as recited in claim 20.

For at least the foregoing reasons, Applicant submits that claim 20 is not anticipated by BURROWS. Accordingly, Applicant respectfully requests that the rejection of claim 20 under 35 U.S.C. § 102(b) based on BURROWS be reconsidered and withdrawn.

Claims 1-2, 5-8, 10, 13, 16-17, 24-29, 31-32, and 34 stand rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over BURROWS in view of WARD. Applicant respectfully traverses this rejection.

Amended independent claim 1 was recited above. BURROWS and WARD, whether taken alone or in any reasonable combination, do not disclose or suggest the combination of features recited in claim 1.

For example, BURROWS and WARD do not disclose or suggest choosing a subset of the plurality of overlapping blocks, where the subset is less than the entirety of the plurality of overlapping blocks, as recited in amended claim 1. The Examiner relies on Fig. 4 of BURROWS for allegedly disclosing this feature (Office Action, p. 8).

Fig. 4 of BURROWS, which is described in col. 7, line 41 to col. 9, line 32,

depicts a block diagram of content attributes generated by the search engine. Fig. 4

shows portions of a page, labeled as 230, 240, 250, and 260, which are detected and

encoded by a parsing module. The parsing module generates attribute values for entire

pages, portions of a page, fields, or individual words and the parser stores these attribute

values as searchable metawords. Fingerprint 255 of Fig. 4 can be one of these

metawords. BURROWS specifically discloses that the fingerprint 255 can be produced

by applying one-way polynomial functions to the digitized contents of the document (col.

8, lines 16-23 of BURROWS).

The Examiner apparently relies on portions 230, 240, 250, and 260 of page 200 in

Fig. 4 as allegedly corresponding to the plurality of overlapping blocks, as recited in

claim 1, and relies on Fig. 4 as allegedly corresponding to choosing a subset of the

plurality of overlapping blocks, as recited in claim 1. However, Fig. 4 of BURROWS

does not disclose or suggest choosing a subset of the set of portions of the pages 230,

240, 250, and 260, where the subset is less than the entirety of the set, as would be

required by amended claim 1, based on the Examiner's interpretation of BURROWS.

Thus, BURROWS in no way discloses choosing a subset of a plurality of overlapping

blocks obtained by sampling a document, where the subset is less than the entirety of the

plurality of overlapping blocks, as recited in amended claim 1. Moreover, Applicant

submits that BURROWS does not disclose or suggest that the different portions of page

200 are overlapping blocks.

Furthermore, BURROWS and WARD do not disclose or suggest compacting the

subset of the plurality of overlapping blocks to obtain the representation of the document,

as also recited in claim 1. The Examiner relies on Fig. 5 of BURROWS for allegedly

disclosing this feature (Office Action, p. 8). The Examiner disagrees with the Examiner interpretation of BURROWS.

Fig. 5 of BURROWS, described in col. 9, lines 33-41, shows a view of the words and metawords produced by the parsing module. The parsing module produces a sequence of pairs in a collating order according to the location of the words of various pages. BURROWS does not disclose or suggest anywhere that this sequence of pairs in a collating order is a representation of the document. Therefore, this section of BURROWS does not disclose or suggest compacting the subset of the plurality of overlapping blocks to obtain the representation of the document, as recited in amended claim 1.

WARD does not overcome the deficiencies of BURROWS set forth above with respect to claim 1.

In the Response to Argument section of the Office Action, the Examiner alleged that choosing a subset of elements in a set can include choosing all the elements (Office Action, p. 6). Amended claim 1 recites choosing a subset of a plurality of overlapping blocks obtained by sampling a document, where the subset is less than the entirety of the plurality of overlapping blocks. Therefore, this argument is moot.

The Examiner alleges that the parsing module 30 of BURROWS produces a sequence of pairs according to the location of words in returned documents, and therefore corresponds to compacting the subset of a plurality of overlapping blocks (Office Action, p. 6). As pointed out above, BORROWS does not disclose or suggest choosing a subset of a plurality of overlapping blocks obtained by sampling a document, where the subset is less than the entirety of the plurality of overlapping blocks. Therefore, BURROWS

cannot disclose or suggest compacting the subset of the plurality of overlapping blocks.

For at least the foregoing reasons, Applicant submits that claim 1 is patentable over BURROWS and WARD, whether taken alone or in any reasonable combination. Accordingly, Applicant respectfully requests that the rejection of claim 1 under 35 U.S.C. § 103(a) based on BURROWS and WARD be reconsidered and withdrawn.

Claims 2, 5-8, and 10 depend from claim 1. Therefore, these claims are patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 1. Accordingly, Applicant respectfully requests that the rejection of claims 2, 5-8, and 10 under 35 U.S.C. § 103(a) based on BURROWS and WARD be reconsidered and withdrawn. Moreover, these claims are patentable over BURROWS and WARD for reasons of their own.

For example, claim 2 recites that compacting the subset of the plurality of overlapping blocks includes setting bits in the representation of the document based on the subset of the plurality of overlapping blocks. BURROWS and WARD do not disclose or suggest this feature.

The Examiner relies on Fig. 5 of BURROWS for allegedly disclosing this feature (Office Action, p. 8). Fig. 5 of BURROWS, described in col. 9, lines 33-41, shows a view of the words and metawords produced by the parsing module. The parsing module produces a sequence of pairs in a collating order according to the location of the words.

Neither this figure nor the description thereof discloses or suggests that compacting the subset of the plurality of overlapping blocks includes setting bits in the representation of the document based on the subset of the plurality of overlapping blocks, as recited in claim 2.

WARD does not overcome the deficiencies of BURROWS set forth above with respect to claim 2.

For at least these additional reasons, Applicant submits that claim 2 is patentable over BURROWS and WARD, whether taken alone or in any reasonable combination.

Independent claim 13 is directed to a method for generating a representation of a document. The method includes sampling the document to obtain a plurality of overlapping samples, selecting a predetermined number of the plurality of overlapping samples as those of the samples corresponding to a predetermined number of smallest samples or a predetermined number of largest samples, and setting bits in the representation of the document based on the selected predetermined number of the samples. BURROWS and WARD do not disclose or suggest this combination of features.

The Examiner did not address the features of claim 13 in the Office Action. Instead, the Examiner addressed the features of claim 1 (Office Action, p. 11). Claim 1 does not, for example, recite selecting a predetermined number of the plurality of overlapping samples as those of the samples corresponding to a predetermined number of smallest samples or a predetermined number of largest samples, as recited in claim 13. The Examiner does not address this feature. Thus, a *prima facie* case of obviousness has not been established with respect to claim 13. If this rejection is maintained, Applicant again respectfully requests that the Examiner specifically address the features recited in claim 13.

Nevertheless, Applicant submits that BURROWS and WARD do not disclose or suggest selecting a predetermined number of the plurality of overlapping samples as

those of the samples corresponding to a predetermined number of smallest samples or a predetermined number of largest samples, as recited in claim 13.

For at least the foregoing reasons, Applicant submits that claim 13 is patentable over BURROWS and WARD, whether taken alone or in any reasonable combination. Accordingly, Applicant respectfully requests that the rejection of claim 13 under 35 U.S.C. § 103(a) based on BURROWS and WARD be reconsidered and withdrawn.

Claims 16-17 depend from claim 13. Therefore, these claims are patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 13. Accordingly, Applicant respectfully requests that the rejection of claims 16-17 under 35 U.S.C. § 103(a) based on BURROWS and WARD be reconsidered and withdrawn.

Claims 24 and 25 depend from claim 20. The disclosure of WARD does not remedy the deficiencies of BURROWS set forth above with respect to claim 20. Therefore, these claims are patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 20. Accordingly, Applicant respectfully requests that the rejection of claims 24-25 under 35 U.S.C. § 103(a) based on BURROWS and WARD be reconsidered and withdrawn.

Independent claims 26 and 34 recite features which are similar to, yet possibly of different scope than, the features recited above with respect to claim 1. Therefore, these claims are patentable over BURROWS and WARD for at least reasons similar to the reasons set forth above with respect to claim 1. Accordingly, Applicant respectfully requests that the rejection of claims 26 and 34 under 35 U.S.C. § 103(a) based on BURROWS and WARD be reconsidered and withdrawn.

Claims 27-29 depend from claim 26. Therefore, these claims are patentable over

BURROWS and WARD for at least the reasons set forth above with respect to claim 26. Accordingly, Applicant respectfully requests that the rejection of claims 27-29 under 35 U.S.C. § 103(a) based on BURROWS and WARD be reconsidered and withdrawn.

Independent claim 31 recites features which are similar to, yet possibly of different scope than, the features recited above with respect to claim 13. Therefore, this claim is patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 13. Accordingly, Applicant respectfully requests that the rejection of claim 31 under 35 U.S.C. § 103(a) based on BURROWS and WARD be reconsidered and withdrawn.

Claim 32 depends from claim 31. Therefore, this claim is patentable over BURROWS and WARD for at least the reasons set forth above with respect to claim 31. Accordingly, Applicant respectfully requests that the rejection of claim 32 under 35 U.S.C. § 103(a) based on BURROWS and WARD be reconsidered and withdrawn.

Claims 3, 4, 11, 12, 14 and 15 stand rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over BURROWS and WARD and further in view of BRODER. Applicant respectfully traverses this rejection.

Claims 3, 4, 11, and 12 depend from claim 1. BRODER does not overcome the deficiencies of BURROWS and WARD set forth above with respect to claim 1. Therefore, these claims are patentable over BURROWS, WARD, and BRODER, for at least the reasons set forth above with respect to claim 1. Accordingly, Applicant respectfully requests that the rejection of claims 3, 4, 11, and 12 under 35 U.S.C. § 103(a) based on BURROWS, WARD, and BRODER be reconsidered and withdrawn.

Claims 14-15 depend from claim 13. BRODER does not overcome the

deficiencies of BURROWS and WARD set forth above with respect to claim 13.

Therefore, these claims are patentable over BURROWS, WARD, and BRODER, for at

least the reasons set forth above with respect to claim 13. Accordingly, Applicant

respectfully requests that the rejection of claims 14-15 under 35 U.S.C. § 103(a) based on

BURROWS, WARD, and BRODER be reconsidered and withdrawn.

Claims 22-23 stand rejected under 35 U.S.C. § 103(a) as allegedly being

unpatentable over BURROWS in view of CHARIKAR. Applicant respectfully traverses

this rejection.

Claims 22-23 depend from claim 20. CHARIKAR does not overcome the

deficiencies of BURROWS set forth above with respect to claim 20. Therefore, these

claims are patentable over BURROWS and CHARIKAR, for at least the reasons set forth

above with respect to claim 20. Accordingly, Applicant respectfully requests that the

rejection of claims 22-23 under 35 U.S.C. § 103(a) based on BURROWS, WARD, and

CHARIKAR be reconsidered and withdrawn.

Claims 9, 18, 30, and 33 stand rejected under 35 U.S.C. § 103(a) as allegedly

being unpatentable over BURROWS and WARD in view of Official Notice. Applicant

respectfully traverses this rejection.

Claim 9 depends from claim 1. The Examiner's taking of Official Notice does

not overcome the deficiencies of BURROWS and WARD set forth above with respect to

claim 1. Therefore, this claim is patentable over BURROWS, WARD and Official

Notice for at least the reasons set forth above with respect to claim 1. Moreover,

Applicant submits that hashing a subset of a plurality of overlapping blocks, which

includes taking a number of least significant bits of the subset of the plurality of

overlapping blocks, was not well known in the art at the time of Applicant's invention.

Applicant respectfully requests that the Examiner provide a reference that discloses that

hashing a subset of a plurality of overlapping blocks, which includes taking a number of

least significant bits of the subset of the plurality of overlapping blocks, was well known

in the art. For at least the foregoing reasons, Applicant respectfully requests that the

rejection of claim 9 under 35 U.S.C. § 103(a) based on BURROWS, WARD, and Official

Notice be reconsidered and withdrawn.

Claim 18 depends from claim 13. The Examiner's taking of Official Notice does

not overcome the deficiencies of BURROWS and WARD set forth above with respect to

claim 13. Therefore, this claim is patentable over BURROWS, WARD and Official

Notice for at least the reasons set forth above with respect to claim 13. Moreover,

Applicant submits that hashing a predetermined number of the samples, which includes

taking a number of least significant bits of the predetermined number of samples, was not

well known in the art at the time of Applicant's invention. Applicant respectfully

requests that the Examiner provide a reference that discloses that hashing a

predetermined number of the samples, which includes taking a number of least

significant bits of the predetermined number of samples, was well known in the art. For

at least the foregoing reasons, Applicant respectfully requests that the rejection of claim

18 under 35 U.S.C. § 103(a) based on BURROWS, WARD, and Official Notice be

reconsidered and withdrawn.

Claim 30 depends from claim 27. The Examiner's taking of Official Notice does

not overcome the deficiencies of BURROWS and WARD set forth above with respect to

claim 27. Therefore, this claim is patentable over BURROWS, WARD and Official

Notice for at least the reasons set forth above with respect to claim 27. Moreover,

Applicant submits that means for compacting the subset of plurality of overlapping

blocks, which include means for flipping bits in the compact representation that are

addressed by a hashed version of the checksum values, was not well known in the art at

the time of Applicant's invention. Applicant respectfully requests that the Examiner

provide a reference that discloses that means for compacting the subset of plurality of

overlapping blocks, which include means for flipping bits in the compact representation

that are addressed by a hashed version of the checksum values, was well known in the art.

For at least the foregoing reasons, Applicant respectfully requests that the rejection of

claim 30 under 35 U.S.C. § 103(a) based on BURROWS, WARD, and Official Notice be

reconsidered and withdrawn.

Claim 33 depends from claim 31. The Examiner's taking of Official Notice does

not overcome the deficiencies of BURROWS and WARD set forth above with respect to

claim 31. Therefore, this claim is patentable over BURROWS, WARD and Official

Notice for at least the reasons set forth above with respect to claim 31. Moreover,

Applicant submits that hashing a predetermined number of the samples, which includes

taking a number of least significant bits of the predetermined number of samples, was not

well known in the art at the time of Applicant's invention. Applicant respectfully

requests that the Examiner provide a reference that discloses that hashing a

predetermined number of the samples, which includes taking a number of least

significant bits of the predetermined number of samples, was well known in the art. For

at least the foregoing reasons, Applicant respectfully requests that the rejection of claim

33 under 35 U.S.C. § 103(a) based on BURROWS, WARD, and Official Notice be

reconsidered and withdrawn.

Claim 21 stands rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over BURROWS and in view of Official Notice. Applicant respectfully traverses this rejection.

Claim 21 depends from claim 20. The Examiner's taking of Official Notice does not overcome the deficiencies of BURROWS set forth above with respect to claim 20. Therefore, this claim is patentable over BURROWS and Official Notice for at least the reasons set forth above with respect to claim 20. Moreover, Applicant submits that a search engine to return documents to a user as a single link when the documents are determined to correspond to near-duplicate documents, was not well known in the art at the time of Applicant's invention. Applicant respectfully requests that the Examiner provide a reference that discloses that a search engine to return documents to a user as a single link when the documents are determined to correspond to near-duplicate documents, was well known in the art. For at least the foregoing reasons, Applicant respectfully requests that the rejection of claim 21 under 35 U.S.C. § 103(a) based on BURROWS and Official Notice be reconsidered and withdrawn.

In view of the foregoing amendments and remarks, Applicant respectfully requests the Examiner's reconsideration of this application, and the timely allowance of the pending claims.

While the present application is believed to be in condition for allowance, should the Examiner find some issue to remain unresolved, or should any new issues arise that could be eliminated through discussions with Applicant's representative, then the

Examiner is invited to contact the undersigned by telephone to expedite prosecution of

the present application.

To the extent necessary, a petition for an extension of time under 37 CFR § 1.136

is hereby made. Please charge any shortage in fees due in connection with the filing of

this paper, including extension of time fees, to Deposit Account No. 50-1070 and please

credit any excess fees to such deposit account.

Respectfully submitted,

HARRITY SNYDER, L.L.P.

By: <u>/ Viktor Simkovic, Reg. No. 56,012/</u>
Viktor Simkovic
Registration No. 56,012

Date: May 8, 2008

Harrity Snyder, L.L.P.
11350 Random Hills Road
Suite 600
Fairfax, Virginia 22030
Main: (571) 432-0800
Direct: (571) 432-0899
Customer Number: 44989